REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
	New Reprint	-

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Non-Markovian State-Dependent Networks in Critical Loading	W911NF-14-1-0216
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
	611102

6. AUTHORS	5d. PROJECT NUMBER
Chihoon Lee, Anatolii A. Puhalskii	
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES	8. PERFORMING ORGANIZATION REPORT NUMBER
Colorado State University - Ft. Collins Sponsored Programs 2002 Campus Delivery Fort Collins, CO 80523 -2002	

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
U.S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709-2211	ARO
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)
	65380-MA.1

12. DISTRIBUTION AVAILIBILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT

We establish a heavy traffic limit theorem for the queue-length process in a critically loaded single class queueing network with state-dependent arrival and service rates. A distinguishing feature of our model is non-Markovian state dependence. The limit stochastic process is a continuous-path reflected process on the nonnegative orthant. We give an application to a generalized Jackson network with state-dependent rates.

15. SUBJECT TERMS

Diffusion approximation; Non-Markovian networks; State-dependent networks; Weak convergence

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Chihoon Lee
UU	UU	UU	UU		19b. TELEPHONE NUMBER
					970-491-7321

Report Title

Non-Markovian State-Dependent Networks in Critical Loading

ABSTRACT

We establish a heavy traffic limit theorem for the queue-length process in a critically loaded single class queueing network with state-dependent arrival and service rates. A distinguishing feature of our model is non-Markovian state dependence. The limit stochastic process is a continuous-path reflected process on the nonnegative orthant. We give an application to a generalized Jackson network with state-dependent rates.

REPORT DOCUMENTATION PAGE (SF298)
(Continuation Sheet)

Continuation for Block 13

Block 13: Supplementary Note

Approved for public release; distribution is unlimited.

Non-Markovian State-Dependent Networks in Critical Loading

Chihoon Lee[a] & Anatolii A. Puhalskii[bc]

[a] Department of Statistics, Colorado State University, Fort Collins, Colorado, USA

[b] University of Colorado Denver, Denver, Colorado, USA

[c] Institute for Problems in Information Transmission, Moscow, Russia
Published online: 04 Feb 2015.

CrossMark

Click for updates

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis
Taylor & Francis Group

NON-MARKOVIAN STATE-DEPENDENT NETWORKS IN CRITICAL LOADING

Chihoon Lee[1] and Anatolii A. Puhalskii[2,3]

[1]*Department of Statistics, Colorado State University, Fort Collins, Colorado, USA*
[2]*University of Colorado Denver, Denver, Colorado, USA*
[3]*Institute for Problems in Information Transmission, Moscow, Russia*

□   *We establish a heavy traffic limit theorem for the queue-length process in a critically loaded single class queueing network with state-dependent arrival and service rates. A distinguishing feature of our model is non-Markovian state dependence. The limit stochastic process is a continuous-path reflected process on the nonnegative orthant. We give an application to a generalized Jackson network with state-dependent rates.*

1. INTRODUCTION

Queueing systems with arrival and (or) service rates depending on the system's state arise in various application areas, which include manufacturing, storage, service engineering, and communication and computer networks. When human servers are involved, longer queues may lead to customers being discouraged to join the queue or may affect productivity. (We provide more detail in Section 5.) State-dependent features are present in congestion control protocols in communication networks, such as TCP (see Refs.[1,3,11,14,20] and references therein).

In this paper, we consider an open network of single server queues with the arrival and service rates depending on the queue lengths. The network consists of K single-server stations indexed 1 through K. Each station has an infinite capacity buffer and the customers are served according to the first-in-first-out discipline. The arrivals of customers at the stations occur both

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/lstm.

externally, from the outside, and internally, from the other stations. Upon service completion at a station, a customer is either routed to another station or exits the network. Every customer entering the network eventually leaves it. A distinguishing feature of the model is non-Markovian state dependence. More specifically, the number of customers at station i, where $i = 1, 2, \ldots, K$, is governed by the following equations:

$$
\begin{aligned}
Q_i(t) &= Q_i(0) + A_i(t) + B_i(t) - D_i(t), \\
A_i(t) &= N_i^A \left(\int_0^t \lambda_i(Q(s)) \, ds \right), \\
B_i(t) &= \sum_{j=1}^K \Phi_{ji}\big(D_j(t)\big), \\
D_i(t) &= N_i^D \left(\int_0^t \mu_i(Q(s)) 1_{\{Q_i(s)>0\}} ds \right),
\end{aligned}
\tag{1.1}
$$

where $Q(s) = (Q_1(s), \ldots, Q_K(s))$ denotes the vector of the queue lengths at the stations at time s. The quantities $N_i^A(t)$ and $N_i^D(t)$ represent the number of exogenous arrivals and the maximal number of customers that can be served, respectively, at station i by time t under "nominal" conditions, $\Phi_{ji}(m)$ represents the number of customers routed from station j to station i out of the first m customers served at station j, and $\lambda_i(Q(t))$ and $\mu_i(Q(t))$ represent instantaneous exogenous arrival and service rates, respectively, for station i at time t given the queue length vector $Q(t)$. Thus, $A_i(t)$ represents the cumulative number of exogenous arrivals by time t at station i, $D_i(t)$ represents the cumulative number of departures by time t from station i, and $B_i(t)$ represents the cumulative number of customers routed to station i from the other stations by time t. The quantities $Q_j(0)$, $N_i^A(t)$, $N_i^D(t)$, and $\Phi_{ij}(m)$ are referred to as network primitives. Generalized Jackson networks is a special case of (1.1) where the N_i^A and N_i^D are renewal processes, the Φ_{ij} are Bernoulli processes, and $\lambda_i(\cdot) = \mu_i(\cdot) = 1$.

Our goal is to obtain limit theorems in critical loading for the queue length processes akin to diffusion approximation results available for generalized Jackson networks; see Reiman[19]. Such limit theorems are useful to obtain approximations to various quantities of interest. For instance, using the limit stochastic processes, one can approximate, via either numerical or analytical methods, the mean/variance of the queue length or the fraction of idle time for particular service stations (cf. Refs.[11,12]).

The results on the heavy traffic asymptotics for state-dependent rates available in the literature are confined mostly to the case of diffusion limits for Markovian models; see Yamada[21], Mandelbaum and Pats[14], and Chapter 8 of Kushner[11]. Yamada[21] and Mandelbaum and Pats[14] draw on the work of Krichagina[10], who studies a Markovian closed network with

state-dependent rates. They consider a case of the model (1.1), where the primitive arrival and service processes are standard Poisson. Kushner[11] also includes a treatment of that model (see Theorem 2.1, p. 318); however, their basic assumptions are formulated in terms of the conditional distributions of the interarrival (or service) intervals (or the routing), given the "past." Those authors obtain results in which the drift coefficients of the limit diffusion processes are state-dependent and the diffusion coefficients may be either constant or state-dependent, which is determined by the scaling used. Yamada[21] and Kushner[11] assume critical loading and obtain mostly diffusions with state-dependent diffusion coefficients, although Yamada[21] considers an example with a constant diffusion coefficient where the drift has to be linear and Theorem 2.1 on p. 318 of Kushner[11] concerns the case of constant diffusion coefficients. In contrast, Mandelbaum and Pats[14] do not restrict their analysis to critical loading, and their limits have constant diffusion coefficients. Mandelbaum and Pats[14] and Kushner[11] also allow the process of routing the customers inside the network to be state-dependent; however, their reasoning seems to be unsubstantiated, as discussed in Section 2. Section 7 of Yamada[21] is concerned with a non-Markovian case where the processes N_i^A are standard Poisson, the processes N_i^D are renewal processes, and $\mu_i(\cdot) = 1$. It is also mentioned that an extension to the case of renewal arrivals with $\lambda_i(\cdot) = 1$ and standard Poisson processes N_i^D is possible.

The main contribution of this piece of work is incorporating general arrival and service processes. This is achieved by applying an approach different from the one used by Yamada[21], Mandelbaum and Pats[14], and Kushner[11]. The proofs of those authors rely heavily on the martingale weak convergence theory. They are quite involved, on the one hand, and do not seem to be easily extendable to more general arrival and service processes, on the other hand. In our approach, we, in a certain sense, return to the basics and employ ideas that have proved their worth in the setup of generalized Jackson networks. We show that continuity considerations may produce stronger conclusions at less complexity. Our main result states that if the network primitives satisfy certain limit theorems with continuous-path limits, then the multidimensional queue-length processes, when suitably scaled and normalized, converge to a reflected continuous-path process on the nonnegative orthant. If the limits of the primitives are diffusion processes, the limit stochastic process is a reflected diffusion with state-dependent drift coefficients and constant diffusion coefficients. The scaling we use does not capture the case of state-dependent diffusion coefficients. We also give an application to generalized Jackson networks with state-dependent rates, thus providing an extension of Reiman's[19] results and particularise further by looking at a tandem queue of human servers. In addition, we bridge certain gaps in the reasoning of Yamada[21], Mandelbaum and Pats[14], and Kushner[11]. The proofs in Refs.[21,14,11] assume that the functions $\lambda_i(\cdot)$ and $\mu_i(\cdot)$ are bounded and allude to a "truncation argument" for the unbounded

case omitting the details. In particular, existence and uniqueness for (1.1) is not fully addressed. We prove both the existence and uniqueness of a solution to equations (1.1) and the limit theorem under linear growth conditions. A more detailed discussion of earlier results is provided at the end of Section 2.

A different class of results on diffusion approximation concerns queueing systems modeled on the many-server queue with a large number of servers. In such a system the service rate decreases to zero gradually with the number in the system—whereas in the model considered here, it has a jump at zero (see (2.1d))—so the limit process is an unconstrained diffusion; see Mandelbaum, Massey, and Reiman[13], Pang, Talreja, and Whitt[16], and references therein. We do not consider those setups in this paper.

The exposition is organised as follows. In the next section, we state and discuss our main result. The proof is provided in Section 3. In Section 4, an application to state-dependent generalized Jackson networks is presented. Section 5 spells out the process for a tandem queue. The appendix contains a proof of the pathwise queue-length construction underlying the definition of the model.

Some notational conventions are in order. All vectors are understood as column vectors, $|x|$ denotes the Euclidean length of a vector x, its components are denoted by x_i, unless mentioned otherwise, superscript ^{T} is used to denote the transpose, 1_A stands for the indicator function of an event A, δ_{ij} represents Kronecker's delta, $\lfloor a \rfloor$ denotes the integer part of a real number a, $\mathbb{N}$ denotes the set of natural numbers, and $\mathbb{Z}_+$ denotes the set of nonnegative integers. We use $\mathbb{D}([0, \infty), \mathbb{R}^\ell)$ to represent the Skorohod space of right continuous $\mathbb{R}^\ell$-valued functions with left hand limits which is endowed with the Skorohod topology, $\Rightarrow$ represents convergence in distribution of random elements with values in an appropriate metric space; see Billingsley[2] and Ethier and Kurtz[6] for more information. We also recall that a sequence V^n of stochastic processes with trajectories in a Skorohod space is said to be $\mathbb{C}$-tight if the sequence of the laws of the V^n is tight, and if all limit points of the sequence of the laws of the V^n are laws of continuous-path processes (see, e.g., Definition 3.25 and Proposition 3.26 in Chapter VI of Jacod and Shiryaev[9]).

2. THE MAIN RESULT

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space on which all random variables considered in this paper are assumed to be defined. We consider a sequence of networks indexed by n with a similar structure to the one described in the Introduction. For the n-th network and for $i \in \mathbb{K}$, where $\mathbb{K} = \{1, 2, \dots, K\}$, let $A_i^n(t)$ represent the cumulative number of customers that arrive at station i from outside the network, let $B_i^n(t)$ represent the

number of endogeneous arrivals at station i, and let $D_i^n(t)$ represent the cumulative number of customers that are served at station i during the time interval $[0, t]$. Let $\mathcal{J} \subseteq \mathbb{K}$ represent the set of stations with actual exogenous arrivals so that $A_i^n(t) = 0$ if $i \notin \mathcal{J}$. We call $A^n = (A_i^n, i \in \mathbb{K})$ and $D^n = (D_i^n, i \in \mathbb{K})$, where $A_i^n = (A_i^n(t), t \geq 0)$ and $D_i^n = (D_i^n(t), t \geq 0)$, the arrival process and service process for the n-th network, respectively. We associate with the stations of the network the processes $\Phi_i^n = (\Phi_{ij}^n, j \in \mathbb{K})$, $i \in \mathbb{K}$, where $\Phi_{ij}^n = (\Phi_{ij}^n(m), m = 1, 2, \ldots)$, and $\Phi_{ij}^n(m)$ denotes the cumulative number of customers among the first m customers departing station i that go directly to station j. The process $\Phi^n = (\Phi_{ij}^n, i, j \in \mathbb{K})$ is referred to as the routing process. The state of the network at time t is described by $Q^n(t) = (Q_1^n(t), \ldots, Q_K^n(t))$, where $Q_i^n(t)$ represents the number of customers at station i at time t.

Precisely, the model is defined as follows. Let λ_i^n and μ_i^n, where $i \in \mathbb{K}$, be Borel functions mapping $\mathbb{R}_+^K$ to $\mathbb{R}_+$, with $\lambda_i^n(x) = 0$ if $i \notin \mathcal{J}$, and let $\lambda^n = (\lambda_i^n, i \in \mathbb{K})$ and $\mu^n = (\mu_i^n, i \in \mathbb{K})$. These functions have the meaning of state-dependent arrival and service rates. Let $N_i^{A,n} = (N_i^{A,n}(t), t \geq 0)$ and $N_i^{D,n} = (N_i^{D,n}(t), t \geq 0)$ represent nondecreasing $\mathbb{Z}_+$-valued processes with trajectories in the Skorohod space $\mathbb{D}([0, \infty), \mathbb{R})$ such that $N_i^{A,n}(0) = N_i^{D,n}(0) = 0$ and the jumps of $N_i^{D,n}$ are of size one. We define $N_i^{A,n}(t) = \lfloor t \rfloor$ if $i \notin \mathcal{J}$. (The latter is but a convenient convention. Since $\lambda_i^n(x) = 0$ if $i \notin \mathcal{J}$, the process $N_i^{A,n}$ is immaterial, as the equations below show.) Let $(\Phi_{ij}^n(m), m \in \mathbb{N})$ be $\mathbb{Z}_+$-valued nondecreasing processes such that $\sum_{j=1}^K \Phi_{ij}^n(m) = m$ and let $Q_i^n(0)$ be $\mathbb{Z}_+$-valued random variables. The processes A_i^n, B_i^n, D_i^n, and Q_i^n are defined as $\mathbb{Z}_+$-valued processes with trajectories in $\mathbb{D}([0, \infty), \mathbb{R})$ satisfying a.s. the equations

$$Q_i^n(t) = Q_i^n(0) + A_i^n(t) + B_i^n(t) - D_i^n(t), \tag{2.1a}$$

$$A_i^n(t) = N_i^{A,n}\left(\int_0^t \lambda_i^n(Q^n(s)) \, ds\right), \tag{2.1b}$$

$$B_i^n(t) = \sum_{j=1}^K \Phi_{ji}^n\big(D_j^n(t)\big), \tag{2.1c}$$

$$D_i^n(t) = N_i^{D,n}\left(\int_0^t \mu_i^n(Q^n(s)) 1_{\{Q^n(s)>0\}} \, ds\right), \tag{2.1d}$$

where $t \geq 0$ and $i \in \mathbb{K}$. Accordingly, Q^n, A^n, and D^n are random elements of $\mathbb{D}([0, \infty), \mathbb{R}^K)$. (Existence and uniqueness for (2.1a)–(2.1d) are addressed in Lemma 2.1.)

Let $P = (p_{ij}, \, i, j \in \mathbb{K})$ be a substochastic matrix, $R = I - P^T$, where I denotes the $K \times K$-identity matrix, and $p_i = (p_{ij}, \, j \in \mathbb{K})$. We denote

$$\overline{Q}^n(0) = \frac{Q^n(0)}{\sqrt{n}}, \qquad\qquad \overline{N}_i^{A,n}(t) = \frac{N_i^{A,n}(nt) - nt}{\sqrt{n}},$$

$$\overline{N}_i^{D,n}(t) = \frac{N_i^{D,n}(nt) - nt}{\sqrt{n}}, \qquad\qquad \overline{\Phi}_i^n(t) = \frac{\Phi_i^n(\lfloor nt \rfloor) - p_i nt}{\sqrt{n}},$$

$$\overline{N}_i^{A,n} = (\overline{N}_i^{A,n}(t), \, t \geq 0), \qquad\qquad \overline{N}^{A,n} = (\overline{N}_i^{A,n}, \, i \in \mathbb{K}),$$

$$\overline{N}_i^{D,n} = (\overline{N}_i^{D,n}(t), \, t \geq 0), \qquad\qquad \overline{N}^{D,n} = (\overline{N}_i^{D,n}, \, i \in \mathbb{K}),$$

$$\overline{\Phi}_i^n = (\overline{\Phi}_i^n(t), \, t \geq 0), \qquad\qquad \overline{\Phi}^n = (\overline{\Phi}_i^n, \, i \in \mathbb{K}) \, .$$

We will need the following conditions:

(A0) For each $n \in \mathbb{N}$ and each $i \in \mathcal{J}$, $\limsup_{t \to \infty} N_i^{A,n}(t)/t < \infty$ a.s.

(A1) The spectral radius of matrix P is strictly less than 1.

(A2) For all $i \in \mathbb{K}$,

$$\sup_{n \in \mathbb{N}} \sup_{x \in \mathbb{R}_+^K} \frac{\lambda_i^n(nx) + \mu_i^n(nx)}{n(1 + |x|)} < \infty.$$

(A3) For all $i \in \mathbb{K}$, there exist continuous functions $\lambda_i(x)$ and $\mu_i(x)$ such that

$$\frac{\lambda_i^n(nx)}{n} \to \lambda_i(x), \qquad \frac{\mu_i^n(nx)}{n} \to \mu_i(x)$$

uniformly on compact subsets of $\mathbb{R}_+^K$, as $n \to \infty$. Furthermore, for all $x \in \mathbb{R}_+^K$,

$$\lambda(x) - R\mu(x) = 0,$$

where $\lambda(x) = (\lambda_i(x), \, i \in \mathbb{K})$ and $\mu(x) = (\mu_i(x), \, i \in \mathbb{K})$.

(A4)

$$\sup_{n \in \mathbb{N}} \sup_{x \in \mathbb{R}_+^K} \frac{1}{\sqrt{n}(1 + |x|)} |\lambda^n(\sqrt{n}x) - R\mu^n(\sqrt{n}x)| < \infty.$$

(A5) There exists a Lipschitz-continuous function $a(x)$ such that

$$\frac{1}{\sqrt{n}}(\lambda^n(\sqrt{n}x) - R\mu^n(\sqrt{n}x)) \to a(x)$$

as $n \to \infty$ uniformly on compact subsets of $\mathbb{R}_+^K$.

(A6) As $n \to \infty$,

$$(\overline{Q}^n(0), \overline{N}^{A,n}, \overline{N}^{D,n}, \overline{\Phi}^n) \Rightarrow (X_0, W^A, W^D, W^\Phi)$$

where X_0 is a random K-vector, $W^A = (W_i^A, i \in \mathbb{K})$, $W^D = (W_i^D, i \in \mathbb{K})$, and $W^\Phi = (W_i^\Phi, i \in \mathbb{K})$ are continuous-path stochastic processes with trajectories in respective spaces $\mathbb{D}([0, \infty), \mathbb{R}^K)$, $\mathbb{D}([0, \infty), \mathbb{R}^K)$, and $\mathbb{D}([0, \infty), \mathbb{R}^{K \times K})$.

Condition (A0) is needed to ensure the existence of a unique strong solution to the system of equations (2.1a)–(2.1d); see Lemma 2.1. It is certainly fulfilled if $N_i^{A,n}$ is a renewal process and is almost a consequence of condition (A6) in that the latter implies that $\lim_{n \to \infty} N_i^{A,n}(nt)/(nt) = 1$ in probability. Part (A1) is essentially an assumption that the network is open and underlies the existence of a regular Skorohod map associated with the network data asserted in Proposition 2.1. The linear growth conditions (A2) and (A4) are needed to ensure the tightness of certain processes. Conditions (A2) and (A3) appear in Mandelbaum and Pats[14]. The requirement that $\lambda(x) = R\mu(x)$ in (A3) together with condition (A5) defines a *critically loaded heavy traffic* regime. Condition (A6) is an assumption on the primitives. The components of W^A corresponding to $i \notin \mathcal{J}$ vanish. Conditions (A2)–(A5) are fulfilled if the following expansions hold:

$$\lambda^n(x) = n\lambda^{(1)}(x/n) + \sqrt{n}\lambda^{(2)}(x/\sqrt{n}) \text{ and } \mu^n(x)$$
$$= n\mu^{(1)}(x/n) + \sqrt{n}\mu^{(2)}(x/\sqrt{n}), \tag{2.2}$$

where $\lambda^{(1)}$ and $\mu^{(1)}$ are continuous functions satisfying the linear-growth condition such that $\lambda^{(1)}(x) = R\mu^{(1)}(x)$, and $\lambda^{(2)}$ and $\mu^{(2)}$ are Lipschitz-continuous bounded functions. If the functions in (2.2) are constant, then one obtains the standard critical loading condition that $(\lambda^n - R\mu^n)/\sqrt{n} \to \lambda_2 - \mu_2$ as $n \to \infty$; cf. Reiman[19].

Lemma 2.1. *Let condition (A0) hold and $\max_{i \in \mathbb{K}} \sup_{x \in \mathbb{R}_+^K}(\lambda_i^n(x) + \mu_i^n(x))/(1 + |x|) < \infty$. Then, given $Q_i^n(0)$, $N_i^{A,n}$, $N_i^{D,n}$, and Φ_{ij}^n equations (2.1a)–(2.1d) admit a unique strong solution Q^n, which is a $\mathbb{Z}_+^K$-valued stochastic process.*

The proof is provided in the appendix. In order to state the main result, we have to recall some properties of the Skorohod map.

Definition 2.1. *Let $\psi \in \mathbb{D}([0, \infty), \mathbb{R}^K)$ be given with $\psi(0) \in \mathbb{R}_+^K$. Then the pair $(\phi, \eta) \in \mathbb{D}([0, \infty), \mathbb{R}^K) \times \mathbb{D}([0, \infty), \mathbb{R}^K)$ solves the Skorohod problem for ψ with respect to $\mathbb{R}_+^K$ and R if the following hold:*

(i) $\phi(t) = \psi(t) + R\eta(t) \in \mathbb{R}_+^K$, for all $t \geq 0$;

(ii) for $i \in \mathbb{K}$, (a) $\eta_i(0) = 0$, (b) η_i is non-decreasing, and (c) η_i can increase only when ϕ is on the ith face of $\mathbb{R}_+^K$, that is, $\int_0^\infty 1_{\{\phi_i(s) \neq 0\}} d\eta_i(s) = 0$.

Let $\mathbb{D}_{\mathbb{R}_+^K}([0, \infty), \mathbb{R}^K) = \{\psi \in \mathbb{D}([0, \infty), \mathbb{R}^K) : \psi(0) \in \mathbb{R}_+^K\}$. If the Skorohod problem has a unique solution on a domain $D \subset \mathbb{D}_{\mathbb{R}_+^K}([0, \infty), \mathbb{R}^K)$, we define the Skorohod map Γ on D by $\Gamma(\psi) = \phi$. The following result (see Harrison and Reiman[7] and also Dupuis and Ishii[4]) yields Lipschitz continuity of the Skorohod map.

Proposition 2.1. *Under assumption (A1), the Skorohod map Γ is well defined on $\mathbb{D}_{\mathbb{R}_+^K}([0, \infty), \mathbb{R}^K)$ and is Lipschitz continuous in the following sense: There exists a constant $L > 0$ such that for all $T > 0$ and $\psi_1, \psi_2 \in \mathbb{D}_{\mathbb{R}_+^K}([0, \infty), \mathbb{R}^K)$,*

$$\sup_{t \in [0, T]} |\Gamma(\psi_1)(t) - \Gamma(\psi_2)(t)| \leq L \sup_{t \in [0, T]} |\psi_1(t) - \psi_2(t)|.$$

Consequently, both ϕ and η are continuous functions of ψ.

The Lipschitz continuity of the Skorohod map and of the function $a(x)$ imply that the equation

$$X(t) = \Gamma\Big(X_0 + \int_0^\cdot a(X(s)) \, ds + M(\cdot)\Big)(t), \qquad (2.3)$$

where

$$M_i(t) = W_i^A(\lambda_i(0) t) + \sum_{j=1}^K W_{ji}^\Phi(\mu_j(0) t) - \sum_{j=1}^K (\delta_{ij} - p_{ji}) W_j^D(\mu_j(0) t), \qquad (2.4)$$

has a unique strong solution (cf.[4]). For $t \geq 0$ and $i \in \mathbb{K}$, let $X_i^n(t) = Q_i^n(t)/\sqrt{n}$. We also define $X = (X(t), t \geq 0)$ and $X^n = ((X_i^n(t), i = 1, 2, \ldots, K), t \geq 0)$. The proof of the following theorem is given in the next section.

Theorem 2.1. *Let conditions (A0)–(A6) hold. Then $X^n \Rightarrow X$, as $n \to \infty$.*

Most of the results on diffusion approximation in critical loading (see, e.g., Harrison and Reiman[7] and Kushner[11]) formulate the heavy traffic condition in terms of rates that are on order one and then consider scaled processes $Q^n(nt)/\sqrt{n}$. In the scaling here, as in Yamada[21] and Mandelbaum and Pats[14], the time parameter is left unchanged, and the factor of n is absorbed in the arrival and service rates. This is more convenient notationally; however, in the application to generalized Jackson networks in Section 4, we work with the conventional scaling. It may be instructive to note, though, that if one looked for limits for processes $Q^n(nt)/\sqrt{n}$, then the analogues of expansions (2.2) would be

$$\lambda^n(x) = \lambda^{(1)}(x/n) + (1/\sqrt{n})\lambda^{(2)}(x/\sqrt{n}) \quad \text{and}$$
$$\mu^n(x) = \mu^{(1)}(x/n) + (1/\sqrt{n})\mu^{(2)}(x/\sqrt{n}),$$

whereas the assumptions of Yamada[21] would amount to the expansions

$$\lambda^n(x) = \lambda^{(1)}(x/\sqrt{n}) + (1/\sqrt{n})\lambda^{(2)}(x/\sqrt{n}) \quad \text{and}$$
$$\mu^n(x) = \mu^{(1)}(x/\sqrt{n}) + (1/\sqrt{n})\mu^{(2)}(x/\sqrt{n}).$$

Theorems 1 and 2 in Yamada[21] obtain diffusion processes with state-dependent drift and diffusion coefficients as the limits. Theorem 1 concerns the Markovian model. It is required that there be at least one nonzero external arrival process. The arrival and service rates at a station may depend on the queue length at that station only. Theorem 2 concerns a Jackson network with external arrival processes being Poisson processes with state-dependent rates. In the proof of Theorem 2, the process $d_n^j(t)$ claimed to be a locally square integrable martingale on p. 980 does not seem to have the asserted property. The model itself is defined by postulating certain martingale properties of the arrival, service, and customer transfer processes. No description in terms of the primitive processes is provided, nor is the issue of the assumptions being self-consistent addressed. The gap is filled in by Mandelbaum and Pats[14], although the authors admit the proof is missing technical detail; see p. 623 in Mandelbaum and Pats[14]. Mandelbaum and Pats[14] give few details with regard to the existence and uniqueness of a solution to (1.1) under linear growth conditions on the rates. Mandelbaum and Pats[14] and Kushner[11] allow the routing matrix to be state-dependent. Mandelbaum and Pats[14] appeal to Theorem 5.1 and Corollary 5.2 in Dupuis and Ishii[5] to substantiate the existence and uniqueness for the associated Skorohod problem; however, those results assume bounded domains, so they do not apply. The authors' attempt on p. 628 to recast the problem as a time-dependent reflection is unconvincing. Kushner[11], in their proofs of Theorem 1.1 on

p. 309 and Theorem 2.1 on p. 318, relies on their Theorem 5.1 on p. 123 and Theorem 5.2 on p. 124, which, in turn, are based on Theorem 2.2 in Dupuis and Ishii[4]. However, those results pertain to reflection directions that are constant on the faces, so they do not apply to state-dependent reflection directions. Kushner[11] does not address the issue of the model being well defined either. Nor are we convinced by the substantiation of the martingale properties claimed to hold on p. 310. Besides, the hypotheses of Theorem 1.1 on p. 309 and Theorem 2.1 on p. 318 of Kushner[11] are missing the condition of the drift and diffusion coefficients being Lipschitz continuous. The functional central limit theorem on p. 584 in Mandelbaum and Pats[14], for the case of critical loading, is missing conditions (A4) and (A5). The proof given in that paper is carried out for a special case where those conditions are met. On the other hand, the condition that the first moments of the initial queue lengths be finite assumed by Mandelbaum and Pats[14] can be left out.

3. PROOF OF THEOREM 2.1

We assume conditions (A0)–(A6) throughout this section. We introduce the "centred" processes as follows: For $i \in \mathbb{K}$ and $t \geq 0$,

$$M_i^n(t) = M_i^{A,n}(t) + M_i^{B,n}(t) - M_i^{D,n}(t), \tag{3.1a}$$

where

$$M_i^{A,n}(t) = N_i^{A,n}\left(\int_0^t \lambda_i^n(Q^n(s))\,ds\right) - \int_0^t \lambda_i^n(Q^n(s))\,ds, \tag{3.1b}$$

$$M_i^{B,n}(t) = \sum_{j=1}^{K}\left(\Phi_{ji}^n\big(D_j^n(t)\big) - p_{ji}D_j^n(t)\right), \tag{3.1c}$$

and

$$
M_i^{D,n}(t) = N_i^{D,n}\left(\int_0^t \mu_i^n(Q^n(s))1_{\{Q^n(s)>0\}}\,ds\right) - \int_0^t \mu_i^n(Q^n(s))1_{\{Q^n(s)>0\}}\,ds
$$
$$
+ \sum_{j=1}^{K} p_{ji}\left(N_j^{D,n}\left(\int_0^t \mu_j^n(Q^n(s))1_{\{Q_j^n(s)>0\}}\,ds\right)\right.
$$
$$
\left. - \int_0^t \mu_j^n(Q^n(s))1_{\{Q_j^n(s)>0\}}\,ds\right). \tag{3.1d}
$$

We can rewrite the evolution (2.1a) as

$$Q_i^n(t) = Q_i^n(0) + \int_0^t \left[\lambda_i^n(Q^n(s)) + \sum_{j=1}^K p_{ji} \mu_j^n(Q^n(s)) - \mu_i^n(Q^n(s)) \right] ds$$
$$+ M_i^n(t) + [RY^n(t)]_i,$$

where $Y^n(t) = (Y_i^n(t), i \in \mathbb{K})$ and

$$Y_i^n(t) = \int_0^t 1_{\{Q_i^n(s)=0\}} \mu_i^n(Q^n(s))\, ds, \quad i \in \mathbb{K}. \tag{3.2}$$

We note that $(Y_i^n(t), t \geq 0)$ is a continuous-path non-decreasing process with $Y_i^n(0) = 0$, which increases only when $Q_i^n(t) = 0$, i.e., $\int_0^\infty 1_{\{Q_i^n(t)\neq 0\}} dY_i^n(t) = 0$ a.s. Let

$$a^n(x) = \lambda^n(x) - R\mu^n(x). \tag{3.3}$$

Then the state evolution can be expressed succinctly by the following vector equation:

$$Q^n(t) = Q^n(0) + \int_0^t a^n(Q^n(s))\, ds + M^n(t) + RY^n(t), \quad t \geq 0, \tag{3.4}$$

where $M^n(t) = (M_i^n(t), i \in \mathbb{K})$. It can also be described in terms of the Skorohod map:

$$Q^n(t) = \Gamma \left(Q^n(0) + \int_0^\cdot a^n(Q^n(s))\, ds + M^n(\cdot) \right)(t), \quad t \geq 0. \tag{3.5}$$

The following tightness result is essential.

Lemma 3.1. *The sequence of processes $(M^n(t)/\sqrt{n}, t \geq 0)$ is $\mathbb{C}$-tight.*

Proof. Let

$$\theta^n(x) = 1 + \sum_{i=1}^K (\mu_i^n(x) + \lambda_i^n(x)), \quad \hat{\mu}_i^n(x) = \frac{\mu_i^n(x)}{\theta^n(x)}, \quad \hat{\lambda}_i^n(x) = \frac{\lambda_i^n(x)}{\theta^n(x)}, \tag{3.6}$$

and

$$\tau^n(t) = \inf \left\{ s : \int_0^s \theta^n(Q^n(u))\, du > t \right\}. \tag{3.7}$$

We note that $\tau^n(t)$ is finite-valued, strictly increasing, absolutely continuous with respect to Lebesgue measure, $d\tau^n(t)/dt = 1/\theta^n(Q^n(\tau^n(t))$ for almost all t, and $\tau^n(t) \to \infty$ as $t \to \infty$; cf. p. 307 in Ethier and Kurtz[6]. Substitution into (2.1a)–(2.1d) shows that the process $\hat{Q}^n = (\hat{Q}^n(t), t \geq 0)$ defined by $\hat{Q}^n(t) = Q^n(\tau^n(t))$ satisfies a.s. the equations

$$
\begin{aligned}
\hat{Q}_i^n(t) &= \hat{Q}_i^n(0) + N_i^{A,n}\left(\int_0^t \hat{\lambda}_i^n(\hat{Q}^n(s))\,ds\right) \\
&\quad + \sum_{j=1}^K \Phi_{ji}^n\left(N_j^{D,n}\left(\int_0^t \hat{\mu}_j^n(\hat{Q}^n(s))1_{\{\hat{Q}_j^n(s)>0\}}\,ds\right)\right) \\
&\quad - N_i^{D,n}\left(\int_0^t \hat{\mu}_i^n(\hat{Q}^n(s))1_{\{\hat{Q}_i^n(s)>0\}}\,ds\right), \quad t \geq 0.
\end{aligned}
\tag{3.8}
$$

Also, $\hat{Q}^n(0) = Q^n(0)$. Let $\hat{\tau}^n(t) = \inf\{s : \tau^n(s) > t\}$ represent the inverse of τ^n. We have that

$$
Q^n(t) = \hat{Q}^n(\hat{\tau}^n(t)).
\tag{3.9}
$$

One can also see that

$$
\hat{\tau}^n(t) = \inf\left\{s : \int_0^s \frac{1}{\theta^n(\hat{Q}^n(u))}\,du > t\right\}.
\tag{3.10}
$$

By (3.8), on noting that $\hat{\lambda}_i^n(x) \leq 1$,

$$
\sum_{i=1}^K \hat{Q}_i^n(t) \leq \sum_{i=1}^K \hat{Q}_i^n(0) + \sum_{i=1}^K N_i^{A,n}(t).
\tag{3.11}
$$

By (3.11) and (A6),

$$
\lim_{r\to\infty} \limsup_{n\to\infty} \mathbf{P}\left(\sup_{s\leq t}\sum_{i=1}^K \frac{1}{n}\hat{Q}_i^n(ns) > r\right) = 0.
\tag{3.12}
$$

By (A2), (3.6), (3.10), and (3.11), for some $H > 0$,

$$
\mathbf{P}(\hat{\tau}^n(t) > nT) \leq \mathbf{P}\left(\int_0^T \frac{n}{H(n + \sum_{i=1}^K \hat{Q}_i^n(0) + \sum_{i=1}^K N_i^{A,n}(nu))}\,du \leq t\right).
\tag{3.13}
$$

Since $\sum_{i=1}^{K} \hat{Q}_i^n(0)/n \to 0$ and $\sum_{i=1}^{K} N_i^{A,n}(nu)/n \to Ku$ in probability as $n \to \infty$, the right-hand side of (3.13) goes to zero for $T > e^{KHt}$. Since by (3.9),

$$\mathbf{P}\left(\sup_{s \leq t} \sum_{i=1}^{K} \frac{1}{n} Q_i^n(s) > r\right) \leq \mathbf{P}(\hat{\tau}^n(t) > n e^{KHt}) + \mathbf{P}\left(\sup_{s \leq e^{KHt}} \sum_{i=1}^{K} \frac{1}{n} \hat{Q}_i^n(ns) > r\right),$$

we conclude by (3.12) that

$$\lim_{r \to \infty} \limsup_{n \to \infty} \mathbf{P}\left(\sup_{s \leq t} \sum_{i=1}^{K} \frac{1}{n} Q_i^n(s) > r\right) = 0. \qquad (3.14)$$

It follows by (A2) that, for $i \in \mathbb{K}$,

$$\lim_{r \to \infty} \limsup_{n \to \infty} \mathbf{P}\left(\int_0^t \left(\frac{1}{n} \lambda_i^n(Q_i^n(s)) + \frac{1}{n} \mu_i^n(Q_i^n(s))\right) ds > r\right) = 0 \qquad (3.15)$$

and that, for $\delta > 0, \epsilon > 0, T > 0$,

$$\lim_{\delta \to 0} \limsup_{n \to \infty} \mathbf{P}\left(\sup_{t \in [0,T]} \int_t^{t+\delta} \left(\frac{1}{n} \lambda_i^n(Q_i^n(s)) + \frac{1}{n} \mu_i^n(Q_i^n(s))\right) ds > \epsilon\right) = 0. \qquad (3.16)$$

By (3.1b), we have that, for $\gamma > 0, \delta > 0, \epsilon > 0, T > 0$, and $r > 0$,

$$\mathbf{P}\left(\sup_{\substack{s,t \in [0,T]: \\ |s-t| \leq \delta}} \left|\frac{1}{\sqrt{n}} M_i^{A,n}(t) - \frac{1}{\sqrt{n}} M_i^{A,n}(s)\right| > \gamma\right)$$

$$\leq \mathbf{P}\left(\int_0^T \frac{1}{n} \lambda_i^n(Q_i^n(s)) \, ds > r\right) + \mathbf{P}\left(\sup_{t \in [0,T]} \int_t^{t+\delta} \frac{1}{n} \lambda_i^n(Q_i^n(s)) \, ds > \epsilon\right)$$

$$+ \mathbf{P}\left(\sup_{\substack{s,t \in [0,r]: \\ |s-t| \leq \epsilon}} \left|\overline{N}_i^{A,n}(t) - \overline{N}_i^{A,n}(s)\right| > \gamma\right).$$

By (A6), the process W^A being continuous, (3.15), and (3.16),

$$
\lim_{\delta \to 0} \limsup_{n \to \infty} \mathbf{P} \left(\sup_{\substack{s, t \in [0, T]: \\ |s-t| \leq \delta}} \left| \frac{1}{\sqrt{n}} M_i^{A,n}(t) - \frac{1}{\sqrt{n}} M_i^{A,n}(s) \right| > \gamma \right) = 0.
$$

Hence, the sequences of processes $(M_i^{A,n}(t)/\sqrt{n},\ t \geq 0)$ are $\mathbb{C}$-tight. A similar argument shows that the sequences of processes $(M_i^{D,n}(t)/\sqrt{n},\ t \geq 0)$ and $(M_i^{B,n}(t)/\sqrt{n},\ t \geq 0)$ are $\mathbb{C}$-tight, so, by (3.1a), the sequence of processes $(M^n(t)/\sqrt{n},\ t \geq 0)$ is $\mathbb{C}$-tight. $\qquad\square$

We identify now the limit points in distribution of $\overline{M}^n = (M^n(t)/\sqrt{n},\ t \geq 0)$.

Lemma 3.2. *The sequence of processes $\overline{M}^n$ converges in distribution, as $n \to \infty$, to* M.

Proof. From Lemma 3.1, $M^n(t)/n \to 0$ in probability uniformly over bounded intervals. By (A2) and (3.3), for some $H > 0$, for all n and x, $|a^n(nx)| \leq Hn(1 + |x|)$. By (3.14), the sequence of processes $(\int_0^t (1/n) a^n(Q^n(s))\, ds,\ t \geq 0)$ is $\mathbb{C}$-tight. By (3.5), (A6), Prohorov's theorem, and the continuity of the Skorohod map, the sequence of processes $(Q^n(t)/n,\ t \geq 0)$ is $\mathbb{C}$-tight. If $(q(t),\ t \geq 0)$ is a limit point in distribution, then, by continuity,

$$
q(t) = \Gamma \left(\int_0^{\cdot} (\lambda(q(s)) - R\mu(q(s)))\, ds \right)(t).
$$

Since by (A3), $\lambda(x) - R\mu(x) = 0$, we must have that $q(t) = 0$, which implies that the sequence $Q_i^n(t)/n$ tends to zero as $n \to \infty$ in probability uniformly on bounded intervals. By (A3) and (3.3), $\int_0^t (1/n) a^n(Q^n(s))\, ds \to 0$ in probability. Since by (3.4) and R being nonsingular, Y^n is a continuous function of $(Q^n(0) + \int_0^t a^n(Q^n(s))\, ds + M^n(t),\ t \geq 0)$, we have that $Y^n(t)/n \to 0$ in probability uniformly over bounded intervals, so by (3.2), for $i \in \mathbb{K}$,

$$
\frac{1}{n} \int_0^t \mu_i^n(Q^n(s)) 1_{\{Q_i^n(s)=0\}}\, ds \to 0 \quad \text{in probability as} \quad n \to \infty. \tag{3.17}
$$

We also have by (A3) and the convergence of $Q^n(t)/n$ to the zero function that

$$\frac{1}{n} \int_0^t \lambda_i^n(Q^n(s)) \, ds \to \lambda_i(0) t \quad \text{in probability as} \quad n \to \infty \tag{3.18a}$$

and

$$\frac{1}{n} \int_0^t \mu_i^n(Q^n(s)) \, ds \to \mu_i(0) t \quad \text{in probability as} \quad n \to \infty. \tag{3.18b}$$

Since by (A6), $N_i^{D,n}(nt)/n \to t$ in probability uniformly over bounded intervals as $n \to \infty$, by (2.1d), (3.17), and (3.18b),

$$\frac{D_i^n(t)}{n} \to \mu_i(0) t \quad \text{in probability as} \quad n \to \infty. \tag{3.19}$$

On recalling the definitions in (3.1b)–(3.1d), we conclude from the convergences in (A6), (3.18a), (3.18b), and (3.19) that the processes $(M^{A,n}/\sqrt{n}, M^{B,n}/\sqrt{n}, M^{D,n}/\sqrt{n})$ jointly converge in distribution to (M^A, M^B, M^D), where $M_i^A(t) = W_i^A(\lambda_i(0)t)$, $M_i^B(t) = \sum_{j=1}^K W_{ji}^\Phi(\mu_j(0)t)$, $M_i^D(t) = \sum_{j=1}^K (\delta_{ij} - p_{ji}) W_j^D(\mu_j(0)t)$. Therefore, by (3.1a) and (2.4), the processes $\overline{M}^n$ converge in distribution to M. □

Proof of Theorem 2.1. We note that by (3.5),

$$X^n(t) = \Gamma \left(X^n(0) + \int_0^{\cdot} \frac{1}{\sqrt{n}} a^n(\sqrt{n} X^n(s)) \, ds + \overline{M}^n(\cdot) \right)(t), \quad t \geq 0. \tag{3.20}$$

By the Lipschitz continuity of the Skorohod map, (3.3), and (A4), for $T > 0$ and suitable $H > 0$ and $L > 0$,

$$\sup_{t \in [0,T]} |X^n(t)| \leq L|X^n(0)| + L \sup_{t \in [0,T]} \int_0^t \frac{1}{\sqrt{n}} |a^n(\sqrt{n} X^n(s))| ds$$

$$+ L \frac{1}{\sqrt{n}} \sup_{t \in [0,T]} |M^n(t)|$$

$$\leq L|X^n(0)| + LH \int_0^T \left(1 + \sup_{s \leq t} |X^n(s)| \right) dt$$

$$+ L \frac{1}{\sqrt{n}} \sup_{t \in [0,T]} |M^n(t)|.$$

Gronwall's inequality, the convergence of the $X^n(0)$ in (A6), and Lemma 3.2 yield

$$\lim_{r \to \infty} \limsup_{n \to \infty} \mathbf{P} \left(\sup_{t \in [0, T]} |X^n(t)| > r \right) = 0,$$

so, by (3.3) and (A5), the sequence of processes $(\int_0^t a^n(\sqrt{n} X^n(s))/\sqrt{n} \, ds, \, t \geq 0)$ is $\mathbb{C}$-tight.

By (3.20), the convergence of the $X^n(0)$, Lemma 3.2, (A5), Prohorov's theorem, and the continuity of the Skorohod map, the sequence of processes $(X^n(t), \, t \geq 0)$ is $\mathbb{C}$-tight and every limit point $(\tilde{X}(t), \, t \geq 0)$ for convergence in distribution satisfies the equation

$$\tilde{X}(t) = \Gamma \left(X(0) + \int_0^\cdot a(\tilde{X}(s)) \, ds + M(\cdot) \right)(t), \quad t \geq 0.$$

The uniqueness of a solution to the Skorohod problem implies that $\tilde{X}(t) = X(t)$. $\qquad \square$

4. GENERALIZED JACKSON NETWORKS WITH STATE-DEPENDENT RATES

In this section, we consider an application to generalized Jackson networks in conventional scaling. We assume as given mutually independent sequences of i.i.d. nonnegative random variables $\{u_j^i(n), i \geq 1\}$ and $\{v_k^i(n), i \geq 1\}$ for $j \in \mathcal{J} \subseteq \mathbb{K}$ and $k \in \mathbb{K}$. For the nth network, the random variable $u_j^i(n)$ represents the ith exogenous interarrival time at station j, while $v_k^i(n)$ represents the ith service time at station k. The variables p_{ij} represent the probabilities of a customer leaving station i being routed directly to station j, which are held constant. The routing decisions, interarrival and service times, and the initial queue length vector are mutually independent.

We define

$$\lambda_j^n = \left(\mathbf{E}[u_j^1(n)] \right)^{-1} > 0, \quad a_j^n = \mathbf{Var}\left(u_j^1(n)\right) \geq 0, \quad j \in \mathcal{J}, \quad \text{and}$$

$$\mu_k^n = \left(\mathbf{E}[v_k^1(n)] \right)^{-1} > 0, \quad s_k^n = \mathbf{Var}\left(v_k^1(n)\right) \geq 0, \quad k \in \mathbb{K},$$

with all these quantities assumed finite and the set $\mathcal{J}$ assumed nonempty. It is convenient to let $\lambda_j^n = 1$ and $a_j^n = 0$ for $j \notin \mathcal{J}$.

Let $\hat{N}_j^{\hat{A}, n}(t) = \max\{i' : \sum_{i=1}^{i'} u_j^i(n) \leq t\}$ for $j \in \mathcal{J}$ and $\hat{N}_k^{\hat{D}, n}(t) = \max\{i' : \sum_{i=1}^{i'} v_k^i(n) \leq t\}$ for $k \in \mathbb{K}$. We may interpret the process $(\hat{N}_j^{\hat{A}, n}(t), \, t \geq 0)$ as a nominal arrival process and the random variables $v_k^i(n)$

as the amounts of work needed to serve the customers. Suppose that arrivals are speeded up (or delayed) by a function $\hat{\lambda}_i^n(x)$, where $i \in \mathcal{J}$, and the service is performed at rate $\hat{\mu}_k^n(x)$, where $k \in \mathbb{K}$, when the queue length vector is x. As in Section 3, we let $\hat{N}_i^{\hat{A},n}(t) = \lfloor t \rfloor$ and $\hat{\lambda}_i^n(x) = 0$ for $i \notin \mathcal{J}$. In analogy with (2.1a)–(2.1d) the queue lengths at the stations at time t, which we denote by $\hat{Q}_i^n(t)$, are assumed to satisfy the equations

$$\hat{Q}_i^n(t) = \hat{Q}_i^n(0) + \hat{A}_i^n(t) + \hat{B}_i^n(t) - \hat{D}_i^n(t),$$

$$\hat{A}_i^n(t) = \hat{N}_i^{\hat{A},n}\left(\int_0^t \hat{\lambda}_i^n(\hat{Q}^n(s))\,ds\right),$$

$$\hat{B}_i^n(t) = \sum_{j=1}^K \hat{\Phi}_{ji}^n\big(\hat{D}_j^n(t)\big),$$

$$\hat{D}_i^n(t) = \hat{N}_i^{\hat{D},n}\left(\int_0^t \hat{\mu}_i^n(\hat{Q}^n(s))1_{\{\hat{Q}_i^n(s)>0\}}\,ds\right),$$

where

$$\hat{\Phi}_{ji}^n(m) = \sum_{l=1}^m \chi_{ji}^n(l),$$

with $\{(\chi_{ji}^n(l), i = 1, 2, \ldots, K), l = 1, 2, \ldots\}$ being Bernoulli random variables that are mutually independent for different j and l and are such that $\mathbf{P}(\chi_{ji}^n(l) = 1) = p_{ji}$.

If we introduce the random variables $Q_i^n(t) = \hat{Q}_i^n(nt)$, $A_i^n(t) = \hat{A}_i^n(nt)$, $B_i^n(t) = \hat{B}_i^n(nt)$, $D_i^n(t) = \hat{D}_i^n(nt)$, $N_i^{A,n}(t) = \hat{N}_i^{\hat{A},n}(t/\lambda_i^n)$, $N_i^{D,n}(t) = \hat{N}_i^{\hat{D},n}(t/\mu_i^n)$, and $\Phi_{ji}^n(m) = \hat{\Phi}_{ji}^n(m)$, and functions $\lambda_i^n(x) = n\lambda_i^n\hat{\lambda}_i^n(x)$ and $\mu_i^n(x) = n\mu_i^n\hat{\mu}_i^n(x)$, then we can see that they satisfy equations (2.1a)–(2.1d). Condition (A0) holds as $N_i^{A,n}(t)/t \to 1$ and $N_i^{D,n}(t)/t \to 1$ a.s. as $t \to \infty$.

Assume that $\hat{Q}^n(0)/\sqrt{n} \Rightarrow X_0$, and for $k \in \mathbb{K}$, $j \in \mathcal{J}$,

$$\lambda_j^n \to \lambda_j, \quad a_j^n \to a_j,$$

$$\mu_k^n \to \mu_k, \quad s_k^n \to s_k,$$

as $n \to \infty$, and also that

$$\max_{k \in \mathbb{K}} \sup_{n \geq 1} \mathbf{E}\big(v_k^1(n)\big)^{2+\epsilon} + \max_{j \in \mathcal{J}} \sup_{n \geq 1} \mathbf{E}\big(u_j^1(n)\big)^{2+\epsilon} < \infty \quad \text{for some} \quad \epsilon > 0.$$

Then condition (A6) holds with $W_j^A = \sqrt{a_j} \lambda_j B_j^A$ for $j \in \mathcal{J}$, $W_j^A(t) = 0$ for $j \notin \mathcal{J}$, $W_k^D = \sqrt{s_k} \mu_k B_k^D$ for $k \in \mathbb{K}$, and with W_i^Φ being a K-dimensional Brownian motion with covariance matrix $\mathbf{E} W_{ik}^\Phi(t) W_{ij}^\Phi(t) = (p_{ij} \delta_{jk} - p_{ij} p_{ik}) t$, where B_j^A and B_k^D are standard Brownian motions and the processes B_j^A, B_k^D, and W_i^Φ are mutually independent.

Let us assume that the following versions of conditions (A2)–(A5) hold:

$\widehat{(A2)}$ For each $i \in \mathbb{K}$,

$$\sup_{n \in \mathbb{N}} \sup_{x \in \mathbb{R}_+^K} \left(\frac{\hat{\lambda}_i^n(nx)}{1 + |x|} + \frac{\hat{\mu}_i^n(nx)}{1 + |x|} \right) < \infty.$$

$\widehat{(A3)}$ There exist continuous functions $\hat{\lambda}_i(x)$ and $\hat{\mu}_i(x)$ such that

$$\hat{\lambda}_i^n(nx) \to \hat{\lambda}_i(x), \quad \hat{\mu}_i^n(nx) \to \hat{\mu}_i(x)$$

uniformly on compact subsets of $\mathbb{R}_+^K$, as $n \to \infty$. Furthermore, for $x \in \mathbb{R}_+^K$,

$$\overline{\lambda}(x) - R\overline{\mu}(x) = 0,$$

where $\overline{\lambda}_i(x) = \lambda_i \hat{\lambda}_i(x)$ and $\overline{\mu}_i(x) = \mu_i \hat{\mu}_i(x)$,

$\widehat{(A4)}$

$$\sup_{n \in \mathbb{N}} \sup_{x \in \mathbb{R}_+^K} \sqrt{n} \frac{|\overline{\lambda}^n(\sqrt{n}x) - R\overline{\mu}^n(\sqrt{n}x)|}{1 + |x|} < \infty.$$

$\widehat{(A5)}$ There exists a Lipschitz-continuous function $\hat{a}(x)$ such that

$$\sqrt{n}(\overline{\lambda}^n(\sqrt{n}x) - R\overline{\mu}^n(\sqrt{n}x)) \to \hat{a}(x)$$

as $n \to \infty$ uniformly on compact subsets of $\mathbb{R}_+^K$, where $\overline{\lambda}_i^n(x) = \lambda_i^n \hat{\lambda}_i^n(x)$ and $\overline{\mu}_i^n(x) = \mu_i^n \hat{\mu}_i^n(x)$.

Then the process M in (2.3) and (2.4) is a K-dimensional Brownian motion with covariance matrix $\mathcal{A}$ that has entries for $i \in \mathbb{K}$,

$$\mathcal{A}_{ii} = \hat{\lambda}_i(0) \lambda_i^3 a_i + \hat{\mu}_i(0) \mu_i^3 s_i (1 - 2p_{ii})$$
$$+ \sum_{j=1}^K \hat{\mu}_j(0) \mu_j p_{ji} (1 - p_{ji} + p_{ji} \mu_j^2 s_j),$$

and for $1 \leq i < j \leq K$,

$$\mathcal{A}_{ij} = - \left[\hat{\mu}_i(0) \mu_i^3 s_i p_{ij} + \hat{\mu}_j(0) \mu_j^3 s_j p_{ji} + \sum_{k=1}^{K} \hat{\mu}_k(0) \mu_k p_{ki} p_{kj} (1 - \mu_k^2 s_k) \right].$$

An application of Theorem 2.1 yields the following result.

Corollary 4.1. *If, in addition to the assumed hypotheses, condition (A1) holds, then the processes $(\hat{Q}^n(nt)/\sqrt{n}, \ t \geq 0)$ converge in distribution to the process $(X(t), \ t \geq 0)$ that satisfies the equation*

$$X(t) = \Gamma \left(X_0 + \int_0^{\cdot} \hat{a}(X(s)) \, ds + \mathcal{A}^{1/2} B(\cdot) \right)(t),$$

where $B(\cdot)$ is a K-*dimensional standard Brownian motion.*

Remark 4.1. The conditions on the asymptotics of the arrival and service rates essentially boil down to the assumptions that the following expansions hold: $\overline{\lambda}^n(x) = \overline{\lambda}^{(1)}(x/n) + \overline{\lambda}^{(2)}(x/\sqrt{n})/\sqrt{n}$ and $\overline{\mu}^n(x) = \overline{\mu}^{(1)}(x/n) + \overline{\mu}^{(2)}(x/\sqrt{n})/\sqrt{n}$ with suitable functions $\overline{\lambda}^{(1)}, \overline{\lambda}^{(2)}, \overline{\mu}^{(1)},$ and $\overline{\mu}^{(2)}$.

Remark 4.2. If, in addition, the assumption of unit rates is made, that is, $\hat{\lambda}_j^n(x) = 1$ for $j \in \mathcal{J}$ and $\hat{\mu}_k^n(x) = 1$ for $k \in \mathbb{K}$, then the limit process is a *K*-dimensional reflected Brownian motion on the positive orthant with infinitesimal drift $\hat{a}(0)$ and covariance matrix $\mathcal{A}$, and the reflection matrix $R = I - P^T$, as in Theorem 1 of Reiman[19].

Remark 4.3. In order to extend applicability, one may consider independent sequences of weakly dependent random variables $\{u_j^i(n), i \geq 1\}$, $\{v_k^i(n), i \geq 1\}$ for $j \in \mathcal{J} \subseteq \mathbb{K}$ and $k \in \mathbb{K}$. Under suitable moment and mixing conditions which imply the invariance principle (cf. Herrndorf[8], Peligrad[17], Jacod and Shiryaev[9]), Corollary 4.1 continues to hold.

5. A CASE STUDY: A TANDEM QUEUE OF HUMAN SERVERS

In production systems with human servers productivity may depend on the workload. With low workload, the productivity may be low because of "laziness," whereas when the workload is too high, the productivity may be adversely affected by the associated "stress"; see van Ooijen and Bertrand[15] and references therein. (More generally, the Yerkes-Dodson law has it that the relation between performance and arousal is given by an inverted U-shaped curve.) Motivated by empirical evidence, van Ooijen and Bertrand[15] use a triangle-shaped function to express the dependence of productivity on
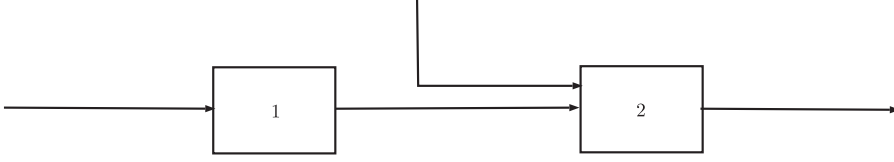
FIGURE 1 The tandem queue.

the workload and investigate the effect of controlling the arrival rate. We will adopt a similar model. We consider two service stations in series (see Figure 1).

Customers who depart from station 1 are routed to station 2 and leave the tandem queue upon completion of service at station 2. Exogenous arrivals occur at both stations. In the nth tandem queue the service rate at station i when the queue length at that station equals x_i is represented by $\mu_i^n(x_i)$, where $i = 1, 2$, and the compression of the interarrival times between the exogenous arrivals is represented by $\lambda_i^n(x_1, x_2)$. The iid service requirements of individual jobs at station i are assumed to have mean $1/\mu_i$ and variance s_i; hence, the service completion rate at station i is $\mu_i \mu_i^n(x_i)$. The iid nominal interarrival times have mean $1/\lambda_i$ and variance a_i so that the actual arrival rate is $\lambda_i \lambda_i^n(x_1, x_2)$.

We assume further that

$$\mu_i^n(x_i) = \tilde{\mu}_{1i}\left(\frac{x_i}{n}\right) + \frac{1}{\sqrt{n}} \tilde{\mu}_{2i}\left(\frac{x_i}{\sqrt{n}}\right),$$

$$\lambda_i^n(x_1, x_2) = \tilde{\lambda}_{1i}\left(\frac{x_1}{n}, \frac{x_2}{n}\right) + \frac{1}{\sqrt{n}} \tilde{\lambda}_{2i}\left(\frac{x_1}{\sqrt{n}}, \frac{x_2}{\sqrt{n}}\right).$$

The functions $\tilde{\mu}_{1i}$ and $\tilde{\mu}_{2i}$ have triangle-shaped graphs as depicted in Figure 2. The exogenous arrival rates are controlled in order to ensure high utilization of the servers, on the one hand, and not to increase the waiting time excessively, on the other hand, for which purpose the critical loading condition is maintained:

$$\begin{pmatrix} \lambda_1 \tilde{\lambda}_{11}(x_1, x_2) \\ \lambda_2 \tilde{\lambda}_{12}(x_1, x_2) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} \mu_1 \tilde{\mu}_{11}(x_1) \\ \mu_2 \tilde{\mu}_{12}(x_2) \end{pmatrix}.$$

In order for these equations to have a nonnegative solution for $\tilde{\lambda}_{12}(x_1, x_2)$, we need to require that $\min_{x_2} \mu_{12} \tilde{\mu}_{12}(x_2) \geq \max_{x_1} \mu_{11} \tilde{\mu}_{11}(x_1)$.

Let $Q_i^n(t)$ denote the queue length at station i at time t. Assuming that $Q_i^n(0) = 0$, we have, by Corollary 4.1, that the processes
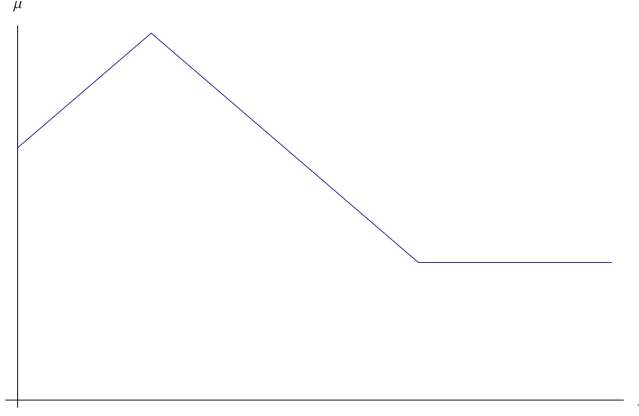
FIGURE 2 Dependence of the service rate on the queue length.

$(Q_1^n(nt)/\sqrt{n}, Q_2^n(nt)/\sqrt{n})$ converge in distribution to the $\mathbb{R}_+^2$-valued process

$$\begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix} = \int_0^t \left(\begin{pmatrix} \lambda_1 \tilde{\lambda}_{21}(X_1(s), X_2(s)) \\ \lambda_2 \tilde{\lambda}_{22}(X_1(s), X_2(s)) \end{pmatrix} - \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} \mu_1 \tilde{\mu}_{21}(X_1(s)) \\ \mu_2 \tilde{\mu}_{22}(X_2(s)) \end{pmatrix} \right) ds$$
$$+ \mathcal{A}^{1/2} \begin{pmatrix} B_1(t) \\ B_2(t) \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} Y_1(t) \\ Y_2(t) \end{pmatrix},$$

where $B_1(t)$ and $B_2(t)$ are independent standard Brownian motions, $Y_1(t)$ and $Y_2(t)$ are nondecreasing continuous path processes such that $Y_i(0) = 0$ and $\int_0^t X_i(s) \, dY_i(s) = 0$, for $i = 1, 2$, and the 2×2 matrix $\mathcal{A}$ has entries

$$\mathcal{A}_{11} = \tilde{\lambda}_{11} \lambda_1^3 a_1 + \tilde{\mu}_{11} \mu_1^3 s_1, \quad \mathcal{A}_{22} = \tilde{\lambda}_{12} \lambda_2^3 a_2 + \tilde{\mu}_{12} \mu_2^3 s_2 + \tilde{\mu}_{11} \mu_1^3 s_1,$$
$$\mathcal{A}_{21} = \mathcal{A}_{12} = -\tilde{\mu}_{11} \mu_1^3 s_1.$$

APPENDIX

Proof of Lemma 2.1. The proof is an adaptation of the one in Puhalskii and Simon[18] (Lemma 2.1) and employs the approach of Ethier and Kurtz[6] (Theorem 4.1, p. 327). We saw in the proof of Lemma 3.1 that if $Q^n(t)$ satisfies (2.1a)–(2.1d), then $\hat{Q}^n(t) = Q^n(\tau^n(t))$, where $\tau^n(t)$ is defined by (3.6) and (3.7), satisfies (3.8). We now show that if $\hat{Q}^n(t)$ satisfies (3.8), then $\hat{Q}^n(\hat{\tau}^n(t))$, where $\hat{\tau}^n(t)$ is defined by (3.10), satisfies (2.1a)–(2.1d). The random variable $\hat{\tau}^n(t)$ is well defined for all t a.s. because, by hypotheses, for

a suitable constant L^n, $\theta^n(x) \le L^n(1 + |x|)$ so that, by (3.11),

$$\int_0^s \frac{1}{\theta^n(\hat{Q}^n(u))} du \ge \frac{1}{L^n} \int_0^s \frac{1}{1 + \sum_{i=1}^K \hat{Q}_i^n(u)} du$$

$$\ge \frac{1}{L^n} \int_0^s \frac{1}{1 + \sum_{i=1}^K \hat{Q}_i^n(0) + \sum_{i=1}^K N_i^{A,n}(u)} du. \quad \text{(A1)}$$

Since $\limsup_{t \to \infty} N_i^{A,n}(t)/t < \infty$ a.s., the rightmost integral in (A1) tends to infinity as $s \to \infty$ a.s., hence, so does the leftmost integral, which proves the claim. In addition, $\hat{\tau}^n(t)$ is absolutely continuous, strictly increasing, $d\hat{\tau}^n(t)/dt = \theta^n(\hat{Q}^n(\hat{\tau}^n(t)))$ a.e., and $\hat{\tau}^n(t) \to \infty$ as $t \to \infty$ a.s. In particular, $\int_0^t \theta^n(\hat{Q}^n(\hat{\tau}^n(s))) \, ds < \infty$. Substitution in (3.8) shows that the random variables $\hat{Q}_i^n(\hat{\tau}^n(t))$ for $i \in \mathbb{K}$ satisfy the equations

$$\hat{Q}_i^n(\hat{\tau}^n(t)) = \hat{Q}_i^n(0) + N_i^{A,n}\Big(\int_0^t \lambda_i^n(\hat{Q}^n(\hat{\tau}^n(s))) \, ds\Big)$$

$$+ \sum_{j=1}^K \Phi_{ji}^n\Big(N_j^{D,n}\Big(\int_0^t \mu_j^n(\hat{Q}^n(\hat{\tau}^n(s))) 1_{\{\hat{Q}_j^n(\hat{\tau}^n(s)) > 0\}} ds\Big)\Big)$$

$$- N_i^{D,n}\Big(\int_0^t \mu_i^n(\hat{Q}^n(\hat{\tau}^n(s))) 1_{\{\hat{Q}_i^n(\hat{\tau}^n(s)) > 0\}} ds\Big), \quad t \ge 0.$$

Therefore, existence and uniqueness for (2.1a)–(2.1d) holds if and only if existence and uniqueness holds for (3.8). The existence and uniqueness for (3.8) follows by recursion on the jump times of $\hat{Q}^n$. More specifically, we define the processes $\hat{Q}^{n,\ell} = (\hat{Q}^{n,\ell}(t), \, t \ge 0)$ with $\hat{Q}^{n,\ell}(t) = (\hat{Q}_i^{n,\ell}(t), \, i \in \mathbb{K})$ by $\hat{Q}_i^{n,0}(t) = \hat{Q}_i^n(0)$ and, for $\ell = 1, 2, \ldots,$ by

$$\hat{Q}_i^{n,\ell}(t) = \Big(\hat{Q}_i^n(0) + N_i^{A,n}\Big(\int_0^t \hat{\lambda}_i^n(\hat{Q}^{n,\ell-1}(s)) \, ds\Big)$$

$$+ \sum_{j=1}^K \Phi_{ji}^n\Big(N_j^{D,n}\Big(\int_0^t \hat{\mu}_j^n(\hat{Q}^{n,\ell-1}(s)) 1_{\{\hat{Q}_j^{n,\ell-1}(s) > 0\}} ds\Big)\Big)$$

$$- N_i^{D,n}\Big(\int_0^t \hat{\mu}_i^n(\hat{Q}^{n,\ell-1}(s)) 1_{\{\hat{Q}_i^{n,\ell-1}(s) > 0\}} ds\Big)\Big)^+. \quad \text{(A2)}$$

The processes $\hat{Q}^{n,\ell}$ are well defined, are $\mathbb{Z}_+^K$-valued, and have right continuous trajectories with left-hand limits. Let $\tau^{n,\ell}$ represent the time epoch of the ℓth jump of $\hat{Q}^{n,\ell}$ with $\tau^{n,0} = 0$. By (A2), $\hat{Q}^{n,1}(t) = \hat{Q}^{n,0}(0)$ if $t < \tau^{n,1}$. Therefore, $\hat{Q}^{n,2}(t) = \hat{Q}^{n,1}(t)$ if $t < \tau^{n,1}$. It follows that $(\hat{Q}^{n,1}(t), \, t \ge 0)$

and $(\hat{Q}^{n,2}(t),\ t \geq 0)$ experience the first jump at the same time epoch and the jump size is the same for both processes, so $\tau^{n,1} < \tau^{n,2}$ and $\hat{Q}^{n,2}(t) = \hat{Q}^{n,1}(t \wedge \tau^{n,1})$ for $t < \tau^{n,2}$. We define $\hat{Q}^{n}(t) = \hat{Q}^{n,0}(0)$ for $t < \tau^{n,1}$ and $\hat{Q}^{n}(t) = \hat{Q}^{n,1}(t)$ for $\tau^{n,1} \leq t < \tau^{n,2}$. Since the processes $N_i^{D,n}$ have unit jumps, we have that (3.8) holds for $t < \tau^{n,2}$. By induction, for arbitrary $\ell \in \mathbb{N}$, we obtain that $\tau^{n,\ell} < \tau^{n,\ell+1}$ and $\hat{Q}^{n,\ell+1}(t) = \hat{Q}^{n,\ell}(t \wedge \tau^{n,\ell})$ for $t < \tau^{n,\ell+1}$. We let $\hat{Q}^{n}(t) = \hat{Q}^{n,\ell}(t)$ for $\tau^{n,\ell} \leq t < \tau^{n,\ell+1}$. The process $\hat{Q}^{n}$ is defined consistently and (3.8) holds for $t \in \cup_{\ell=1}^{\infty}[\tau^{n,\ell-1}, \tau^{n,\ell})$. If $\tau^{n,\ell+1} = \infty$ for some ℓ, then we let $\hat{Q}^{n}(t) = \hat{Q}^{n,\ell}(t)$ for all $t \geq \tau^{n,\ell}$.

Suppose that $\tau^{n,\ell} < \infty$ for all ℓ. Then $\hat{Q}^{n}(t)$ has been defined for all $t < \tau^{n,\infty} = \lim_{\ell \to \infty} \tau^{n,\ell}$ and satisfies (3.8) for these values of t. We now show that $\tau^{n,\infty} = \infty$. The set of the time epochs of the jumps of $\hat{Q}^{n}$ is a subset of the set of the time epochs of the jumps of the process $\tilde{Q}^{n} = (\tilde{Q}^{n}(t),\ t \geq 0)$, where

$$\tilde{Q}^{n}(t) = \sum_{i=1}^{K}\left(N_i^{A,n}\Big(\int_0^t \hat{\lambda}_i^n(\hat{Q}^n(s))\,ds\Big) + N_i^{D,n}\Big(\int_0^t \hat{\mu}_i^n(\hat{Q}^n(s))1_{\{\hat{Q}^n(s)>0\}}ds\Big)\right).$$

Since the process $\hat{Q}^{n}$ has infinitely many jumps, so does the process $\tilde{Q}^{n}$. Since $\hat{\lambda}_i^n(x) \leq 1$ and $\mu_i^n(x) \leq 1$, the lengths of time between the jumps of $\tilde{Q}^{n}$ are not less than the lengths of time between the corresponding jumps of the process $\left(\sum_{i=1}^{K} N_i^{A,n}(t) + \sum_{i=1}^{K} N_i^{D,n}(t),\ t \geq 0\right)$. The process $\left(\sum_{i=1}^{K} N_i^{A,n}(t) + \sum_{i=1}^{K} N_i^{D,n}(t),\ t \geq 0\right)$ having infinitely many jumps and being finite for all t, the time epochs of the jumps of that process tend to infinity as the jump numbers tend to infinity. Thus, $\tau^{n,\infty} = \infty$ a.s.

The provided construction shows that $\hat{Q}^{n}$ is a suitably measurable function of $\hat{Q}^{n}(0)$, $N^{A,n}$, $N^{D,n}$, and Φ^n, so it is a strong solution. The uniqueness is proved similarly. More specifically, suppose $\check{Q}^{n}$ is another solution to (3.8). Let $\check{\tau}^{n,\ell}$ represents the time epoch of the ℓth jump of $\check{Q}^{n}$. We also let $\check{\tau}^{n,0} = 0$. Suppose that $\check{Q}^{n}(t) = \hat{Q}^{n}(t)$ for $t \leq \tau^{n,\ell}$ which is true for $\ell = 0$. Then $\check{Q}^{n}(t) = \hat{Q}^{n}(t) = \hat{Q}^{n}(\tau^{n,\ell})$ for $t < \tau^{n,\ell+1} \wedge \check{\tau}^{n,\ell+1}$. Hence, the integrals from 0 to $\tau^{n,\ell+1} \wedge \check{\tau}^{n,\ell+1}$ on the right of (3.8) are the same for $\check{Q}^{n}$ and $\hat{Q}^{n}$, which implies that $\check{Q}^{n}(\tau^{n,\ell+1} \wedge \check{\tau}^{n,\ell+1}) = \hat{Q}^{n}(\tau^{n,\ell+1} \wedge \check{\tau}^{n,\ell+1})$. Since one of the processes jumps at $\tau^{n,\ell+1} \wedge \check{\tau}^{n,\ell+1}$ the other must jump too, so $\tau^{n,\ell+1} = \check{\tau}^{n,\ell+1}$. $\qquad\square$

FUNDING

REFERENCES

1. Bekker, R. *Queues with State-dependent Rates*. Ph.D. thesis, Technische Universiteit Eindhoven, Eindhoven, 2005.
2. Billingsley, P. *Convergence of Probability Measures*, 2nd edition. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc.: New York, 1999.
3. Chan, W. K.; Schruben, L. W. Mathematical programming representations for state-dependent queues. In S. J. Mason, R. R. Hill, L. Mönch, O. Rose, T. Jefferson, and J. W. Fowler, editors, *Proceedings of the 2008 Winter Simulation Conference*, pages 495–501. IEEE, 2008.
4. Dupuis, P.; Ishii, H. On Lipschitz continuity of the solution mapping to the Skorohod problem, with applications. *Stochastics* **1991**, *35*, 31–62.
5. Dupuis, P.; Ishii, H. SDEs with oblique reflection on nonsmooth domains. *Ann. Probab,* **1993**, *21*, 554–580. Correction: *Ann. Probab,* **2008**, *36*, 1992–1997.
6. Ethier, S. N.; Kurtz, T. G. *Markov Processes. Characterization and Convergence* . Wiley: New York, 1986.
7. Harrison, J. M.; Reiman, M. I. Reflected Brownian motion on an orthant. *Ann. Probab,* **1981**, *9*, 302–308.
8. Herrndorf, N. A functional central limit theorem for weakly dependent sequences of random variables. *Ann. Probab.* **1984**, *12*, 141–153.
9. Jacod, J.; Shiryaev, A. N. *Limit Theorems for Stochastic Processes,* volume 288 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag: Berlin, second edition, 2003.
10. Krichagina, E. V. Asymptotic analysis of queueing networks (martingale approach). *Stochastics Stochastics Rep.* **1992**, *40*, 43–76.
11. Kushner, H. J. *Heavy traffic analysis of controlled queueing and communication networks,* volume 47 of *Applications of Mathematics*. Springer-Verlag: New York, 2001.
12. Kushner, H. J.; Dupuis, P. *Numerical methods for stochastic control problems in continuous time,* volume 24 of *Stochastic Modelling and Applied Probability*, second edition. Springer-Verlag: New York, 2001.
13. Mandelbaum, A. Massey, W. A., Reiman, M. Strong approximations for Markovian service networks. *Queueing Syst.* **1998**, *30*, 149–201.
14. Mandelbaum, A.; Pats, G. State-dependent stochastic networks. I. Approximations and applications with continuous diffusion limits. *Ann. Appl. Probab.* **1998**, *8*, 569–646.
15. van Ooijen, H. P. G.; Bertrand, J. W. M. The effects of a simple arrival rate control policy on throughput and work-in-process in production systems with workload dependent processing rates. *Int. J. Prod. Econ.* **2003**, *85*, 61–68.
16. Pang, G.; Talreja, R.; Whitt, W. Martingale proofs of many-server heavy-traffic limits for Markovian queues. *Probab. Surv.* **2007**, *4*, 193–267.
17. Peligrad, M. Invariance principles under weak dependence. *J. Multivariate Anal.* **1986**, *19*, 299–310.
18. Puhalskii, A.; Simon, B. Discrete evolutionary birth-death processes and their large population limits. *Stochastic Models* **2012**, *28*, 388–412.
19. Reiman, M. I. Open queueing networks in heavy traffic. *Math. Oper. Res.* **1984**, *9*, 441–458.
20. Whitt, W. Queues with service times and interarrival times depending linearly and randomly upon waiting times. *Queueing Syst. Theory Appl.* **1990**, *6*, 335–351.
21. Yamada, K. Diffusion approximation for open state-dependent queueing networks in the heavy traffic situation. *Ann. Appl. Probab.* **1995**, *5*, 958–982.